Medidas de associação

Regressão Linear Simples

Felipe Figueiredo

UNIAN - Centro Universitário Anhanguera de Niterói

Sumário

Sumário

Modelos estatísticos

Modelos servem para:

- representar de forma simplificada fenômenos, experimentos, dados, etc;
- possibilitar análise em cenários controlados, menos complexos que a realidade;
- extrapolar resultados e conclusões.

Modelos estatísticos

Modelos servem para:

- representar de forma simplificada fenômenos, experimentos, dados, etc;
- possibilitar análise em cenários controlados, menos complexos que a realidade;
- extrapolar resultados e conclusões.

Modelos estatísticos

Modelos servem para:

- representar de forma simplificada fenômenos, experimentos, dados, etc;
- possibilitar análise em cenários controlados, menos complexos que a realidade;
- extrapolar resultados e conclusões.

Modelos estatísticos

Ao ajustar um modelo aos dados, podemos:

- fazer predições dentro do intervalo observado para dados que não foram obtidos (interpolação)
- fazer predições fora do intervalo observado (extrapolação)

Modelos estatísticos

Ao ajustar um modelo aos dados, podemos:

- fazer predições dentro do intervalo observado para dados que não foram obtidos (interpolação)
- fazer predições fora do intervalo observado (extrapolação)

Reta de regressão

Definition

Uma reta de regressão (também chamada de reta de melhor ajuste) é a reta para a qual a soma dos erros quadráticos dos resíduos é o mínimo.

- É a reta que melhor se ajusta aos dados
- Minimiza os resíduos

Reta de regressão

Definition

Uma reta de regressão (também chamada de reta de melhor ajuste) é a reta para a qual a soma dos erros quadráticos dos resíduos é o mínimo.

- É a reta que melhor se ajusta aos dados
- Minimiza os resíduos

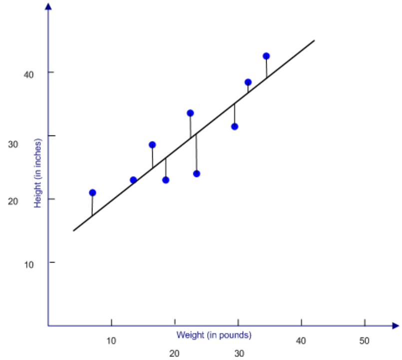
Reta de regressão

Definition

Uma reta de regressão (também chamada de reta de melhor ajuste) é a reta para a qual a soma dos erros quadráticos dos resíduos é o mínimo.

- É a reta que melhor se ajusta aos dados
- Minimiza os resíduos

Resíduos



Definition

Resíduos são a distância entre o dado observado e a reta estimada (modelo).

Elementos da reta de regressão

- Relembrando: a equação de uma reta é definida pela fórmula
$$\hat{y} = ax + b$$
- No caso da reta regressora:
  - y é a variável dependente
  - x é a variável independente
  - a é a inclinação
  - b é o intercepto
- Assim, o objetivo da análise de regressão é encontrar os valores a e b

Elementos da reta de regressão

- Relembrando: a equação de uma reta é definida pela fórmula
$$\hat{y} = ax + b$$
- No caso da reta regressora:
  - y é a variável dependente
  - x é a variável independente
  - a é a inclinação
  - b é o intercepto
- Assim, o objetivo da análise de regressão é encontrar os valores a e b

Elementos da reta de regressão

- Relembrando: a equação de uma reta é definida pela fórmula
$$\hat{y} = ax + b$$
- No caso da reta regressora:
  - y é a variável dependente
  - x é a variável independente
  - a é a inclinação
  - b é o intercepto
- Assim, o objetivo da análise de regressão é encontrar os valores a e b

Elementos da reta de regressão

- Relembrando: a equação de uma reta é definida pela fórmula
$$\hat{y} = ax + b$$
- No caso da reta regressora:
  - y é a variável dependente
  - x é a variável independente
  - a é a inclinação
  - b é o intercepto
- Assim, o objetivo da análise de regressão é encontrar os valores a e b

Elementos da reta de regressão

- Relembrando: a equação de uma reta é definida pela fórmula
$$\hat{y} = ax + b$$
- No caso da reta regressora:
  - y é a variável dependente
  - x é a variável independente
  - a é a inclinação
  - b é o intercepto
- Assim, o objetivo da análise de regressão é encontrar os valores a e b

Elementos da reta de regressão

- Relembrando: a equação de uma reta é definida pela fórmula
$$\hat{y} = ax + b$$
- No caso da reta regressora:
  - y é a variável dependente
  - x é a variável independente
  - a é a inclinação
  - b é o intercepto
- Assim, o objetivo da análise de regressão é encontrar os valores a e b

Elementos da reta de regressão

- Relembrando: a equação de uma reta é definida pela fórmula
$$\hat{y} = ax + b$$
- No caso da reta regressora:
  - y é a variável dependente
  - x é a variável independente
  - a é a inclinação
  - b é o intercepto
- Assim, o objetivo da análise de regressão é encontrar os valores a e b

Análise de Regressão

Para determinar a inclinação e o intercepto, usamos:

- as médias de X e Y
- as variâncias de X e Y
- o coeficiente de correlação r entre X e Y
- o tamanho da amostra n
- ...e algumas operações entre estes termos

Análise de Regressão

Para determinar a inclinação e o intercepto, usamos:

- as médias de X e Y
- as variâncias de X e Y
- o coeficiente de correlação r entre X e Y
- o tamanho da amostra n
- ... e algumas operações entre estes termos

Análise de Regressão

Para determinar a inclinação e o intercepto, usamos:

- as médias de X e Y
- as variâncias de X e Y
- o coeficiente de correlação r entre X e Y
- o tamanho da amostra n
- ...e algumas operações entre estes termos

Análise de Regressão

Para determinar a inclinação e o intercepto, usamos:

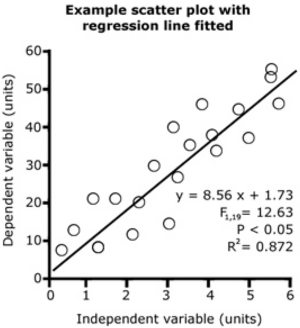
- as médias de X e Y
- as variâncias de X e Y
- o coeficiente de correlação r entre X e Y
- o tamanho da amostra n
- ...e algumas operações entre estes termos

Análise de Regressão

Para determinar a inclinação e o intercepto, usamos:

- as médias de X e Y
- as variâncias de X e Y
- o coeficiente de correlação r entre X e Y
- o tamanho da amostra n
- ... e algumas operações entre estes termos

Análise de Regressão



- A qualidade do ajuste do modelo de regressão é determinado pelo coeficiente de determinação r^2

Sumário

Coeficiente de Determinação r^2

Definition

O coeficiente de determinação r^2 é a relação da variação explicada com a variação total.

$$r^2 = \frac{\text{variação explicada}}{\text{variação total}}$$

- Lembrando: r^2 é o quadrado de r!

Coeficiente de Determinação r^2

Definition

O coeficiente de determinação r^2 é a relação da variação explicada com a variação total.

$$r^2 = \frac{\text{variação explicada}}{\text{variação total}}$$

- Lembrando: r^2 é o quadrado de r!

Coeficiente de Determinação r^2

- Qual é a porcentagem da variação dos dados pode ser explicada pela reta regressora?
- O coeficiente r^2 é a fração da variância que é compartilhada entre X e Y.
- Como r está sempre entre -1 e 1, r^2 está sempre entre 0 e 1.

Coeficiente de Determinação r^2

- Qual é a porcentagem da variação dos dados pode ser explicada pela reta regressora?
- O coeficiente r^2 é a fração da variância que é compartilhada entre X e Y.
- Como r está sempre entre -1 e 1, r^2 está sempre entre 0 e 1.

Coeficiente de Determinação r^2

- Qual é a porcentagem da variação dos dados pode ser explicada pela reta regressora?
- O coeficiente r^2 é a fração da variância que é compartilhada entre X e Y.
- Como r está sempre entre -1 e 1, r^2 está sempre entre 0 e 1.

Coeficiente de Determinação r^2

- Além disso, $r^2 \leq |r|$
- Por que?

Compare os seguintes números entre 0 e 1:

$$\frac{1}{2} \text{ e } \left(\frac{1}{2}\right)^2 = \frac{1}{4} \Rightarrow \frac{1}{4} \leq \frac{1}{2}$$

$$\frac{1}{3} \text{ e } \left(\frac{1}{3}\right)^2 = \frac{1}{9} \Rightarrow \frac{1}{9} \leq \frac{1}{3}$$

Coeficiente de Determinação r^2

- Além disso, $r^2 \leq |r|$
- Por que?

Compare os seguintes números entre 0 e 1:

$$\frac{1}{2} \text{ e } \left(\frac{1}{2}\right)^2 = \frac{1}{4} \Rightarrow \frac{1}{4} \leq \frac{1}{2}$$

$$\frac{1}{3} \text{ e } \left(\frac{1}{3}\right)^2 = \frac{1}{9} \Rightarrow \frac{1}{9} \leq \frac{1}{3}$$

Coeficiente de Determinação r^2

- Além disso, $r^2 \leq |r|$
- Por que?

Compare os seguintes números entre 0 e 1:

$$\frac{1}{2} \text{ e } \left(\frac{1}{2}\right)^2 = \frac{1}{4} \Rightarrow \frac{1}{4} \leq \frac{1}{2}$$

$$\frac{1}{3} \text{ e } \left(\frac{1}{3}\right)^2 = \frac{1}{9} \Rightarrow \frac{1}{9} \leq \frac{1}{3}$$

Interpretação

- Se a correlação é 0, então X e Y não variam juntos (independentes)
- Se a correlação é positiva, então quando uma aumenta, a outra aumenta em proporção direta (linear)
- Se a correlação é negativa, então quando uma aumenta, a outra diminui em proporção inversa (linear)

Interpretação

- Se a correlação é 0, então X e Y não variam juntos (independentes)
- Se a correlação é positiva, então quando uma aumenta, a outra aumenta em proporção direta (linear)
- Se a correlação é negativa, então quando uma aumenta, a outra diminui em proporção inversa (linear)

Interpretação

- Se a correlação é 0, então X e Y não variam juntos (independentes)
- Se a correlação é positiva, então quando uma aumenta, a outra aumenta em proporção direta (linear)
- Se a correlação é negativa, então quando uma aumenta, a outra diminui em proporção inversa (linear)

Cuidado!

- Duas variáveis podem parecer correlacionadas pois são influenciadas por uma terceira variável
- Ex: em alguns países a mortalidade infantil é negativamente correlacionada com o número de telefones per capita
- Mas comprar mais telefones não vai salvar crianças!
- Explicação alternativa: a melhoria da condições financeiras pode afetar ambas as variáveis

Cuidado!

- Duas variáveis podem parecer correlacionadas pois são influenciadas por uma terceira variável
- Ex: em alguns países a mortalidade infantil é negativamente correlacionada com o número de telefones per capita
- Mas comprar mais telefones não vai salvar crianças!
- Explicação alternativa: a melhoria da condições financeiras pode afetar ambas as variáveis

Cuidado!

- Duas variáveis podem parecer correlacionadas pois são influenciadas por uma terceira variável
- Ex: em alguns países a mortalidade infantil é negativamente correlacionada com o número de telefones per capita
- Mas comprar mais telefones não vai salvar crianças!
- Explicação alternativa: a melhoria da condições financeiras pode afetar ambas as variáveis

Cuidado!

- Duas variáveis podem parecer correlacionadas pois são influenciadas por uma terceira variável
- Ex: em alguns países a mortalidade infantil é negativamente correlacionada com o número de telefones per capita
- Mas comprar mais telefones não vai salvar crianças!
- Explicação alternativa: a melhoria da condições financeiras pode afetar ambas as variáveis

Causa x efeito

- Se há uma relação de causalidade entre as duas variáveis, a correlação será não nula (positiva ou negativa)
- Quanto maior for a relação de dependência entre as variáveis, maior será o módulo da correlação.
- Se as variáveis não são relacionadas, a correlação será nula.

Causa x efeito

- Se há uma relação de causalidade entre as duas variáveis, a correlação será não nula (positiva ou negativa)
- Quanto maior for a relação de dependência entre as variáveis, maior será o módulo da correlação.
- Se as variáveis não são relacionadas, a correlação será nula.

Causa x efeito

- Se há uma relação de causalidade entre as duas variáveis, a correlação será não nula (positiva ou negativa)
- Quanto maior for a relação de dependência entre as variáveis, maior será o módulo da correlação.
- Se as variáveis não são relacionadas, a correlação será nula.

Causalidade?

- Mas não podemos inverter a afirmativa lógica do slide anterior!
- Isto é, ao observar uma forte correlação, gostaríamos de concluir que uma variável causa este efeito na outra
- Infelizmente isto não é possível!
- Lembre-se: a significância do teste indica a probabilidade de se cometer um erro do tipo I (falso positivo).

Repita várias vezes mentalmente

Correlação não implica em causalidade.

Causalidade?

- Mas não podemos inverter a afirmativa lógica do slide anterior!
- Isto é, ao observar uma forte correlação, gostaríamos de concluir que uma variável causa este efeito na outra
- Infelizmente isto não é possível!
- Lembre-se: a significância do teste indica a probabilidade de se cometer um erro do tipo I (falso positivo).

Repita várias vezes mentalmente

Correlação não implica em causalidade.

Causalidade?

- Mas não podemos inverter a afirmativa lógica do slide anterior!
- Isto é, ao observar uma forte correlação, gostaríamos de concluir que uma variável causa este efeito na outra
- Infelizmente isto não é possível!
- Lembre-se: a significância do teste indica a probabilidade de se cometer um erro do tipo I (falso positivo).

Repita várias vezes mentalmente

Correlação não implica em causalidade.

Causalidade?

- Mas não podemos inverter a afirmativa lógica do slide anterior!
- Isto é, ao observar uma forte correlação, gostaríamos de concluir que uma variável causa este efeito na outra
- Infelizmente isto não é possível!
- Lembre-se: a significância do teste indica a probabilidade de se cometer um erro do tipo I (falso positivo).

Repita várias vezes mentalmente

Correlação não implica em causalidade.

Causalidade?

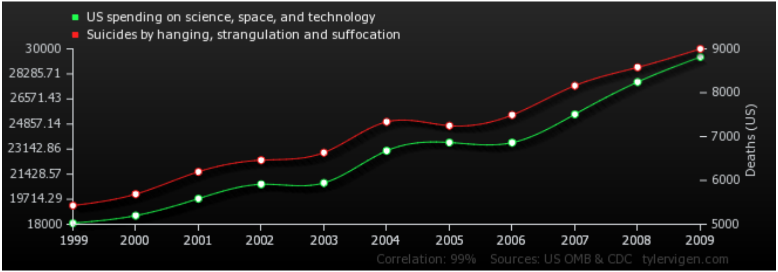
- Mas não podemos inverter a afirmativa lógica do slide anterior!
- Isto é, ao observar uma forte correlação, gostaríamos de concluir que uma variável causa este efeito na outra
- Infelizmente isto não é possível!
- Lembre-se: a significância do teste indica a probabilidade de se cometer um erro do tipo I (falso positivo).

Repita várias vezes mentalmente

Correlação não implica em causalidade.

Exemplo

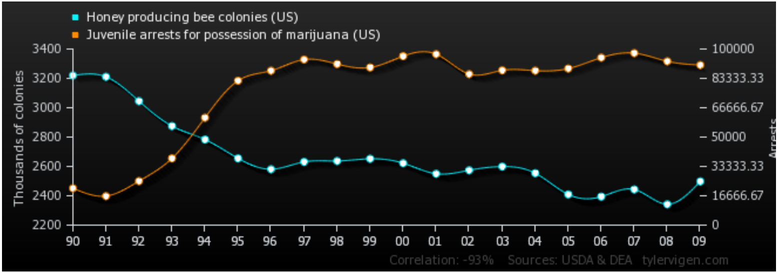
Gasto com C&T (EUA) x Suicídios por enforcamento



Correlação: 0.992082
(Fonte: Spurious correlations)

Exemplo

Produção de mel x Prisões por posse de maconha

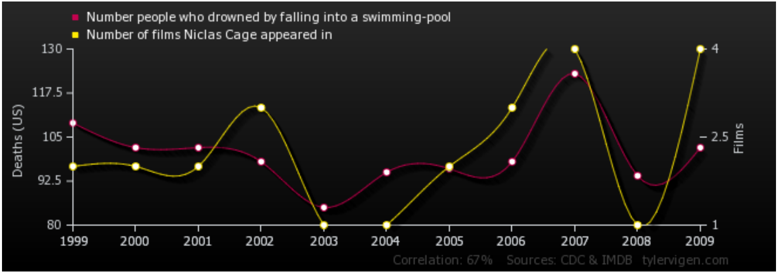


Correlação: -0.933389

(Fonte: Spurious correlations)

Exemplo

Afogamentos em piscina x Filmes com Nicholas Cage



Correlação: 0.666004
(Fonte: Spurious correlations)

Causa e efeito

Ao encontrar uma forte correlação, deve-se sempre se perguntar:

1. Há uma relação direta de causa e efeito entre as variáveis? (X causa Y?)
2. Há uma relação inversa de causa e efeito entre as variáveis? (Y causa X?)
3. É possível que a relação entre as variáveis possa ser causada por uma terceira variável (ou mais) que não foi analisada?
4. É possível que a relação entre duas variáveis seja uma coincidência?

Causa e efeito

Ao encontrar uma forte correlação, deve-se sempre se perguntar:

1. Há uma relação direta de causa e efeito entre as variáveis? (X causa Y?)
2. Há uma relação inversa de causa e efeito entre as variáveis? (Y causa X?)
3. É possível que a relação entre as variáveis possa ser causada por uma terceira variável (ou mais) que não foi analisada?
4. É possível que a relação entre duas variáveis seja uma coincidência?

Causa e efeito

Ao encontrar uma forte correlação, deve-se sempre se perguntar:

1. Há uma relação direta de causa e efeito entre as variáveis? (X causa Y?)
2. Há uma relação inversa de causa e efeito entre as variáveis? (Y causa X?)
3. É possível que a relação entre as variáveis possa ser causada por uma terceira variável (ou mais) que não foi analisada?
4. É possível que a relação entre duas variáveis seja uma coincidência?

Causa e efeito

Ao encontrar uma forte correlação, deve-se sempre se perguntar:

1. Há uma relação direta de causa e efeito entre as variáveis? (X causa Y?)
2. Há uma relação inversa de causa e efeito entre as variáveis? (Y causa X?)
3. É possível que a relação entre as variáveis possa ser causada por uma terceira variável (ou mais) que não foi analisada?
4. É possível que a relação entre duas variáveis seja uma coincidência?

Resumo

- É necessário investigar a relação entre as variáveis!
- O que pode explicar a relação observada?
- Qual proporção (porcentagem) da variabilidade pode ser explicada pelas variáveis analisadas?
- Quão bem a reta regressora se ajusta aos dados?

Resumo

- É necessário investigar a relação entre as variáveis!
- O que pode explicar a relação observada?
- Qual proporção (porcentagem) da variabilidade pode ser explicada pelas variáveis analisadas?
- Quão bem a reta regressora se ajusta aos dados?

Resumo

- É necessário investigar a relação entre as variáveis!
- O que pode explicar a relação observada?
- Qual proporção (porcentagem) da variabilidade pode ser explicada pelas variáveis analisadas?
- Quão bem a reta regressora se ajusta aos dados?

Resumo

- É necessário investigar a relação entre as variáveis!
- O que pode explicar a relação observada?
- Qual proporção (porcentagem) da variabilidade pode ser explicada pelas variáveis analisadas?
- Quão bem a reta regressora se ajusta aos dados?